

On the Role of Diversity in Sparsity Estimation

Galen Reeves and Michael Gastpar[†]

Department of Electrical Engineering and Computer Sciences
University of California, Berkeley

Abstract—A major challenge in sparsity pattern estimation is that small modes are difficult to detect in the presence of noise. This problem is alleviated if one can observe samples from multiple realizations of the nonzero values for the same sparsity pattern. We will refer to this as “diversity”. Diversity comes at a price, however, since each new realization adds new unknown nonzero values, thus increasing uncertainty. In this paper, upper and lower bounds on joint sparsity pattern estimation are derived. These bounds, which improve upon existing results even in the absence of diversity, illustrate key tradeoffs between the number of measurements, the accuracy of estimation, and the diversity. It is shown, for instance, that diversity introduces a tradeoff between the uncertainty in the noise and the uncertainty in the nonzero values. Moreover, it is shown that the optimal amount of diversity significantly improves the behavior of the estimation problem for both optimal and computationally efficient estimators.

I. INTRODUCTION

An extensive amount of recent research in signal processing and statistics has focused on multivariate regression problems with sparsity constraints. One problem of particular interest, known as *sparsity pattern estimation*, is to determine which coefficients are nonzero using a limited number of observations. Remarkably, it has been shown that accurate estimation is possible using a relatively small number of (possibly noisy) linear measurements, provided that the number of nonzero values is relatively small (see e.g. [1]–[3]).

It has also been shown that the presence of additional structure, beyond sparsity, can significantly alter the problem. Various examples include distributed or model-based compressed sensing [4]–[6], estimation from multiple measurement vectors [7], simultaneous sparse approximation [8], model selection [9], union support recovery [10], multi-task learning [11], and estimation of block-sparse signals [12], [13].

In the present paper, we consider a joint sparsity pattern estimation framework motivated in part by the following engineering problem. Suppose that one wishes to estimate the sparsity pattern of an unknown vector and is allowed to take either M noisy linear measurements of the vector itself, or spread the same number measurements amongst multiple vectors with same sparsity pattern as the original vector, but different nonzero values. This type of problem arises, for example, in magnetic resonance imaging where the vectors correspond to images of the same body part (common sparsity pattern) viewed with different contrasting agents (different nonzero values).

[†] Also with the School of Computer and Communication Sciences, EPFL, Lausanne, Switzerland.

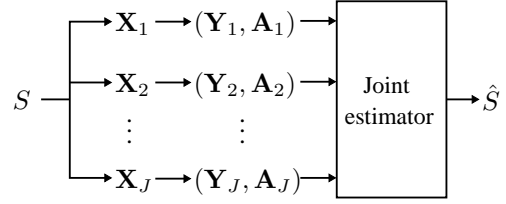


Fig. 1. Illustration of joint sparsity pattern estimation. The vectors \mathbf{X}_j share a common sparsity pattern S but have independent nonzero values. The sparsity pattern S is estimated jointly using measurements vectors \mathbf{Y}_i corresponding to different measurement matrices \mathbf{A}_j .

On one hand, splitting measurements across different vectors increases the number of unknown values, potentially making estimation more difficult. On the other hand, using all measurements on a single vector has the risk that nonzero values with small magnitudes will not be detected. To understand this tradeoff, this paper bounds the accuracy of various estimators for the estimation problem illustrated in Figure 1. We refer to the number of vectors J as the “diversity”.

A. Overview of Contributions

Several key contributions of this paper are the following:

- Our analysis improves upon previous work in the single-vector setting [2], [3] and shows that there exists a sharp divide between knowing almost everything and knowing almost nothing about the sparsity pattern, even in problem regimes where exact recovery is impossible.
- Our bounds are relatively tight for a large range of problem parameters. Unlike bounds based on the restricted isometry property, they apply even when the number of measurements is small relative to the size of the sparsity pattern.
- We show that the right amount of diversity is beneficial, but too much or too little can be detrimental (when the total number of measurements is fixed). Moreover, we show that diversity can significantly reduce the gap in performance between computationally efficient estimators, such as the matched filter or LASSO, and estimators without any computational constraints.

The remainder of the paper is outlined as follows. Theorem 1 gives a sufficient condition for a combinatorial estimator. Theorem 2 gives an information-theoretic necessary condition for any estimator. Theorem 3 gives a necessary and sufficient condition for a two-stage estimation architecture corresponding to either the matched filter, the LASSO, or the MMSE vector estimators, and Theorems 4–6 characterize various

tradeoffs between the diversity, the number of measurements, the SNR, and the accuracy of estimation.

Finally, we note that the joint estimation problem in this paper is closely related to the multiple measurement vector problem [7], except that each vector is measured using a different matrix. Alternatively, our problem is a special case of block-sparsity [12], [13] with a block-sparse measurement matrix. Versions of our bounds for block-sparsity with dense measurement matrices can also be derived.

B. Problem Formulation

Let $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_J \in \mathbb{R}^n$ be a set of jointly random sparse vectors whose nonzero values are indexed by a common sparsity pattern S

$$S = \{i : X_j(i) \neq 0\}, \quad \text{for } j = 1, 2, \dots, J. \quad (1)$$

We assume that S is distributed uniformly over all subsets of $\{1, 2, \dots, n\}$ of size k where k is known, and that the nonzero values are i.i.d. $\mathcal{N}(0, 1)$ ¹.

We consider estimation of S from measurement vectors $\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_J \in \mathbb{R}^m$ of the form

$$\mathbf{Y}_j = \sqrt{\frac{\text{SNR}}{k}} \mathbf{A}_j \mathbf{X}_j + \mathbf{W}_j \quad \text{for } j = 1, 2, \dots, J \quad (2)$$

where each $\mathbf{A}_j \in \mathbb{R}^{m \times n}$ is a known matrix whose elements are i.i.d. $\mathcal{N}(0, 1)$ and $\mathbf{W}_j \sim \mathcal{N}(0, I_{m \times m})$ is unknown noise. The estimation problem is depicted in Figure 1. The accuracy of an estimate \hat{S} is assessed using the (normalized) distortion function

$$d(S, \hat{S}) = \frac{1}{k} \max(|S \setminus \hat{S}|, |\hat{S} \setminus S|) \quad (3)$$

where $|S \setminus \hat{S}|$ and $|\hat{S} \setminus S|$ denote the number of missed detections and false alarms respectively.

Our analysis considers the high dimensional setting where the diversity J is fixed but the vector length n , sparsity k , and number of measurements per vector m tend to infinity. We focus exclusively on the setting of linear sparsity where $k/n \rightarrow \kappa$ for some fixed *sparsity rate* $\kappa \in (0, 1/2)$ and $m/n \rightarrow r$ for some fixed *per-vector sampling rate* $r > 0$. The total number of measurements is given by $M = mJ$, and we use $\rho = Jr$ to denote the total sampling rate. We say that a distortion $\alpha \geq 0$ is *achievable* for an estimator \hat{S} if $\Pr[d(S, \hat{S}) > \alpha] \rightarrow 0$ as $n \rightarrow \infty$. The case $\alpha = 0$ corresponds to exact recovery and the case $\alpha > 0$ corresponds to a constant fraction of errors.

C. Notations

For a matrix A and set of integers S we use $A(S)$ to denote the matrix formed by concatenating the columns of A indexed by S . We use $H_b(p) = -p \log p - (1-p) \log(1-p)$ to denote binary entropy and all logarithms are natural.

¹The results in this paper extend to any i.i.d. distribution with bounded second moment. Due to space constraints, only the Gaussian case is presented.

II. JOINT ESTIMATION BOUNDS

This section gives necessary and sufficient conditions for the joint sparsity pattern estimation problem depicted in Figure 1.

One important property of the estimation problem is the relative size of the smallest nonzero values, averaged across realizations. For a given fraction $\beta \in [0, 1]$, we define random variable

$$P_J^{(n)}(\beta) = \arg \min_{\Delta \subset S : |\Delta| = \alpha k} \frac{1}{J} \sum_{j=1}^J \|\mathbf{X}_j(\Delta)\|^2. \quad (4)$$

By the Glivenko-Cantelli theorem, $P_J^{(n)}(\beta)$ converges almost surely to a nonrandom limit $P_J(\beta)$. We will refer to this limit as the *diversity power*. If the nonzero values are Gaussian, as is assumed in this paper, it can be shown that

$$P_J(\beta) = \int_0^\alpha \xi_J(p) dp \quad (5)$$

where

$$\xi_J(p) = \{t : \mathbb{P}[\frac{1}{J} \chi_J^2 \leq t] = p\} \quad (6)$$

denotes the quantile function of a normalized chi-square random variable with J degrees of freedom.

Another important property is the metric entropy rate (in nats per vector length) of S with respect to our distortion function $d(S, \hat{S})$. In [2], it is shown that this rate is given by

$$R(\kappa, \alpha) = H(\kappa) - \kappa H_b(\alpha) - (1-\kappa) H_b(\frac{\kappa\alpha}{1-\kappa}) \quad (7)$$

for all $\alpha < 1 - \kappa$ and is equal to zero otherwise.

A. Nearest Subspace Upper Bound

We first consider the *nearest subspace* (NS) estimator which is given by

$$\hat{S}^{\text{NS}} = \arg \min_{S : |S|=k} \sum_{j=1}^J \text{dist}(\mathbf{Y}_j, \mathbf{A}_j(S))^2 \quad (8)$$

where $\text{dist}(\mathbf{Y}_j, \mathbf{A}_j(S))$ denotes the euclidean distance between \mathbf{Y}_j and the linear subspace spanned by the columns of $\mathbf{A}_j(S)$. (For the case $J = 1$, this estimator is known variously throughout the literature as ℓ_0 minimization or maximum likelihood estimation.)

Theorem 1. For a given set $(\kappa, \text{SNR}, \rho, J)$, a distortion α is achievable for the nearest subspace estimator if

$$\rho > \kappa J + \max_{\beta \in [\alpha, 1]} \min(E_1(\beta), E_2(\beta)) \quad (9)$$

where

$$E_1(\beta) = \frac{2H_b(\kappa) - 2R(\kappa, \beta) + 2\beta\kappa J \log(5/3)}{\frac{1}{J} \log(1 + \frac{4}{25} J P_J(\beta) \text{SNR})} \quad (10)$$

$$E_2(\beta) = \frac{2H_b(\kappa) - 2R(\kappa, \beta)}{\log(1 + P_1(\beta) \text{SNR}) + 1/(P_1(\beta) \text{SNR}) - 1} \quad (11)$$

with $P_J(\cdot)$ given by (5) and $R(\cdot, \cdot)$ given by (7).

Theorem 1 is a combination of two bounds. The part due to $E_1(\beta)$ determines the scaling behavior at low distortions and low SNR and the part due to $E_2(\beta)$ determines the scaling behavior at high SNR. One important property of $E_1(\beta)$ is that its denominator scales linearly with the effective power of the $P_J(\beta)$ SNR when β is small. As a consequence, Theorem 1 closes a gap in previous bounds for the case $J = 1$ and correctly characterizes the boost in performance due to the diversity when $J > 1$.

B. Optimal Estimation

We next consider an information-theoretic lower bound on the distortion for any estimator. This bound depends on the entropy of the smallest nonzero values. For a given fraction $\beta \in [0, 1]$, we define the *conditional entropy power*

$$\mathcal{N}(\beta) = \frac{1}{2\pi e} \exp \left\{ -2h(U|U^2 \leq \xi_1(\beta)) \right\} \quad (12)$$

where $h(\cdot)$ is differential entropy and $U \sim \mathcal{N}(0, 1)$.

Theorem 2. For a given set $(\kappa, \text{SNR}, \rho, J)$, a distortion α is not achievable for any estimator if

$$\max_{\beta \in [0, 1]} \left\{ R\left(\frac{\beta\kappa}{1-\kappa+\beta\kappa}, \frac{\alpha}{\beta}\right) - J \min(\Lambda_1(\beta), \Lambda_2(\beta)) \right\} > 0 \quad (13)$$

where

$$\Lambda_1(\beta) = \mathcal{V}_1\left(\frac{\rho}{1-\kappa+\beta\kappa}, P_J^2(\beta) \text{SNR}\right) \quad (14)$$

$$\Lambda_2(\beta) = \mathcal{V}_1\left(\frac{\rho}{1-\kappa+\beta\kappa}, \beta^{1-1/J} P_1(\beta^{1/J}) \text{SNR}\right) - \frac{\beta\kappa}{1-\kappa+\beta\kappa} \mathcal{V}_2\left(\frac{\rho}{\beta\kappa}, \beta \mathcal{N}(\beta^{1/J}) \text{SNR}\right) \quad (15)$$

with

$$\mathcal{V}_1(r, \gamma) = \begin{cases} \frac{r}{2} \log(1 + \gamma), & \text{if } r \leq 1 \\ \frac{1}{2} \log(1 + r\gamma), & \text{if } r > 1 \end{cases} \quad (16)$$

$$\mathcal{V}_2(r, \gamma) = \begin{cases} \frac{r}{2} \log(1 + \gamma \Delta(r)), & \text{if } r < 1 \\ \frac{1}{2} \log(1 + r\gamma \Delta(\frac{1}{r})), & \text{if } r > 1 \end{cases} \quad (17)$$

and $\Delta(r) = e^{-1}(1-r)^{1-1/r}$.

Theorem 2 is also a combination of two bounds. The part due to $\Lambda_1(\beta)$ determines the scaling behavior at low distortions and low SNR and the part due to $\Lambda_2(\beta)$ determines the scaling behavior at high SNR. As was the case for the nearest subspace upper bound, this bound is inversely proportional to the effective power $P_J(\beta) \text{SNR}$ when the effective power is small.

III. TWO-STAGE ESTIMATION BOUNDS

This section gives bounds for the two-stage estimation architecture depicted in Figure 2. In the first stage, each vector \mathbf{X}_j is estimated from its measurements \mathbf{Y}_j . In the second stage, the sparsity pattern S is estimated by jointly thresholding estimates $\hat{\mathbf{X}}_1, \hat{\mathbf{X}}_2, \dots, \hat{\mathbf{X}}_J$. One advantage of this architecture is that the estimation in the first stage can be done in parallel. We will see that this architecture can be near optimal in some settings but is highly suboptimal in others.

A. Single-Vector Estimation

Three different estimators are considered: the *matched filter* (MF), the *LASSO*, and the *minimum mean squared error* estimator (MMSE). Recent results have shown that the asymptotic behavior of these estimators can be characterized in terms of an equivalent scalar estimation problem. Since these results correspond to the case $J = 1$, we use the notation \mathbf{X} and \mathbf{Y} and use the per-vector sampling rate r instead of the total sampling rate ρ . Also, we define the sparse Gaussian distribution

$$F_\kappa(x) = \kappa \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du + (1 - \kappa) \mathbf{1}(x \leq 0) \quad (18)$$

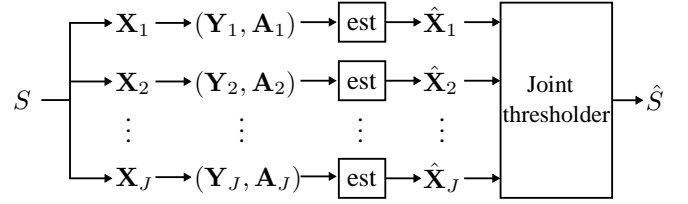


Fig. 2. Illustration of single-vector estimation followed by joint thresholding.

which corresponds to the marginal distribution of $X(i)$.

The first result characterizes the asymptotic behavior of the matched filter which is given by

$$\hat{\mathbf{X}}^{\text{MF}} = \frac{1}{m} \sqrt{\frac{k}{\text{SNR}}} \mathbf{A}^T \mathbf{Y}. \quad (19)$$

To our knowledge, this result was first shown (with convergence in probability) in [3], [14]. Almost sure convergence follows from recent tools developed in [15].

Proposition 1 (Matched Filter). The empirical distribution on the elements of $(\mathbf{X}, \hat{\mathbf{X}}^{\text{MF}})$ converges weakly and almost surely to the distribution on $(X, X + \sigma W)$ where $X \sim F_\kappa$ and $W \sim \mathcal{N}(0, 1)$ are independent and

$$\sigma^2 = \frac{\kappa}{r} \left[\frac{1}{\text{SNR}} + 1 \right]. \quad (20)$$

The next result, due to Donoho et al. [16] and Bayati and Montanari, [15], describes the asymptotic behavior of the LASSO which is given by

$$\hat{\mathbf{X}}^{\text{LASSO}} = \arg \inf_{\mathbf{x} \in \mathbb{R}^n} \frac{1}{2} \|\mathbf{Y} - \sqrt{\frac{\text{SNR}}{k}} \mathbf{A} \mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_1 \quad (21)$$

where $\lambda \geq 0$ is a regularization parameter.

Proposition 2 (LASSO). The empirical distribution on the elements of $(\mathbf{X}, \hat{\mathbf{X}}^{\text{LASSO}})$ converges weakly almost surely to the distribution on $(X, \eta_t(X + \sigma W))$ where $X \sim F_\kappa$ and $W \sim \mathcal{N}(0, 1)$ are independent, $\eta_t(x) = [x - \text{sign}(x)t] \mathbf{1}(|x| > t)$, and σ^2 and t are given by the fixed point equations

$$\sigma^2 = \frac{1}{r} \left[\frac{\kappa}{\text{SNR}} + \mathbb{E}[|X - \eta_t(X + \sigma W)|^2] \right] \quad (22)$$

$$t = \frac{1}{r} \left[\frac{\kappa}{\text{SNR}} \lambda + t \Pr[|X + \sigma W| > t] \right]. \quad (23)$$

The final result, based on the work of Guo and Verdu [17], characterizes the asymptotic behavior of the MMSE which is given by

$$\hat{\mathbf{X}}^{\text{MMSE}} = \mathbb{E}[\mathbf{X}|\mathbf{Y}]. \quad (24)$$

This result depends on a powerful but non-rigorous replica method, and is thus stated as a claim.

Claim 1 (MMSE). The distribution on the elements of $(\mathbf{X}, \hat{\mathbf{X}}^{\text{MMSE}})$ converges weakly in expectation to the distribution on $(X, \mathbb{E}[X|X + \sigma W])$ where $X \sim F_\kappa$ and $W \sim \mathcal{N}(0, 1)$ are independent and σ^2 is given by

$$\sigma^2 = \arg \min_{\sigma^2 \geq 0} \left\{ r \log \sigma^2 + \frac{\kappa}{\text{SNR} \sigma^2} + 2 I(X; X + \sigma W) \right\}. \quad (25)$$

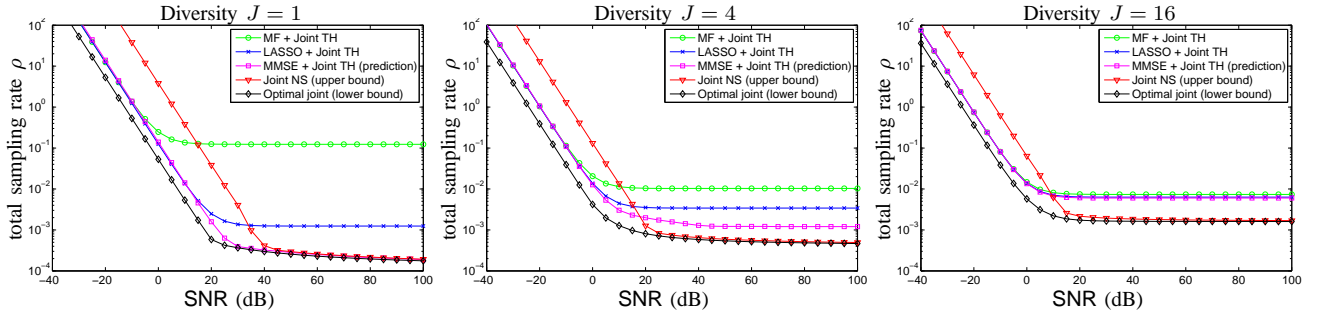


Fig. 3. Bounds on the total sampling rate $\rho = Jr$ as a function of SNR for various J when $\alpha = 0.1$ and $\kappa = 10^{-4}$.

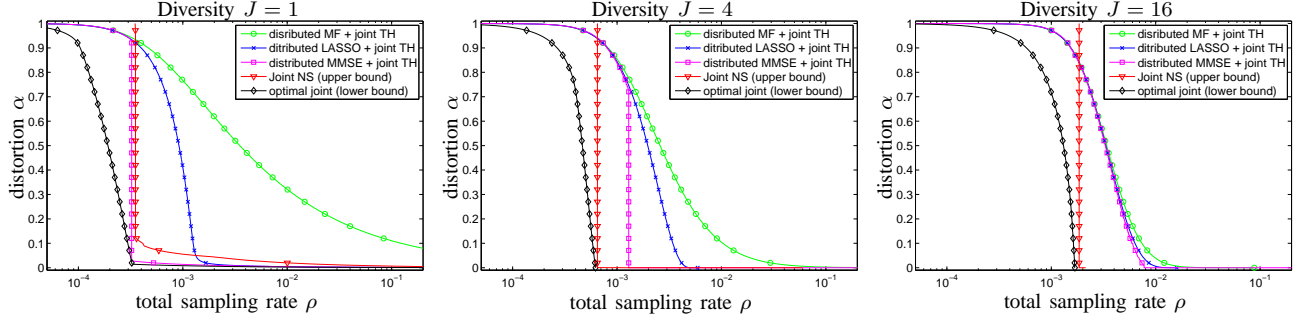


Fig. 4. Bounds on the distortion α as a function of the total sampling rate $\rho = Jr$ for various J when SNR = 40 dB and $\kappa = 10^{-4}$.

B. Thresholding

For the second stage of estimation we consider the *joint thresholding* sparsity pattern estimator given by

$$\hat{S}^{\text{TH}} = \{i : \sum_{j=1}^J \hat{X}_j^2(i) \geq t\} \quad (26)$$

where the threshold $t \geq 0$ is chosen to minimize the expected distortion. Since this estimator evaluates each index $i \in \{1, 2, \dots, n\}$ independently, and since the estimated vectors $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_J$ are conditionally independent given the sparsity pattern S , the distribution on the distortion $d(S, \hat{S}^{\text{TH}})$ can be characterized by joint distribution on $(X_1(1), \hat{X}_1(1))$.

The following result describes the relationship between the distortion α , the diversity J , and the effective noise power σ^2 .

Theorem 3. Suppose that for $j = 1, 2, \dots, J$, the empirical joint distributions on the elements of $(\mathbf{X}_j, \hat{\mathbf{X}}_j)$ converge weakly to one of the scalar distributions corresponding to the matched filter (Proposition 1), the LASSO (Proposition 2), or the MMSE (Claim 1) with noise power σ^2 . Then, a distortion α is achievable for the thresholding estimator (26) if and only if $\sigma^2 \geq \sigma_J^2(\alpha)$ where

$$\sigma_J^2(\alpha) = \frac{\xi_J(\alpha)}{\xi_J(1 - \frac{\alpha\kappa}{1-\kappa}) - \xi_J(\alpha)} \quad (27)$$

with $\xi_J(\alpha)$ given by (6).

Theorem 3 shows that the relationship between α and J is encapsulated by the term $\sigma_J^2(\alpha)$. With a bit of work it can be shown that the numerator and denominator in (27) scale like $\alpha^{-1}P_J(\alpha)$ and $\alpha^{-1}R(\kappa, \alpha)$ respectively when α is small. Thus, plugging $\sigma_J^2(\alpha)$ into the equivalent noise expression of the matched filter given in (20) shows that bounds attained using Theorem 3 have similar low distortion behavior to the bounds in Section II.

One advantageous property of Theorem 3 is that the bounds are exact. As a consequence, these bounds are sometimes lower than the upper bound in Theorem 1, which is loose in general. One shortcoming however, is that the two-stage architecture does not take full advantage of the joint structure during the first stage of estimation. As a consequence, the performance of these estimators can be highly suboptimal, especially at high SNR.

IV. SAMPLING - DIVERSITY TRADEOFF

In this section, we analyze various behaviors of the bounds in Theorems 1, 2, and 3, with an emphasis on the tradeoff provided by the diversity J . The following results characterize the high SNR and low distortion behavior of optimal estimation.

Theorem 4 (High SNR). Let (κ, J, α) be fixed and let $\rho(\text{SNR})$ denote the infimum over sampling rates ρ such that α is achievable for the optimal estimator. Fix any $\epsilon > 0$.

(a) If $\alpha > 0$, then

$$\rho(\text{SNR}) \leq J\kappa + \frac{2H_b(\kappa)(1 + \epsilon)}{\log \text{SNR}} \quad (28)$$

for all SNR large enough.

(b) If $2R(\kappa, \alpha) > J\kappa$, then

$$\rho(\text{SNR}) \geq J\kappa + \frac{2R(\kappa, \alpha)(1 - \epsilon)}{\log \text{SNR}} \quad (29)$$

for all SNR large enough.

Theorem 5 (Low Distortion). Let (κ, J, SNR) be fixed and let $\rho(\alpha)$ denote the infimum over sampling rates ρ such that α is achievable for the optimal estimator. There exist constants $0 < C^- \leq C^+ < \infty$ such that

$$C^- \left(\frac{1}{\alpha}\right)^{2/J} \log\left(\frac{1}{\alpha}\right) \leq \rho(\alpha) \leq C^+ \left(\frac{1}{\alpha}\right)^{2/J} \log\left(\frac{1}{\alpha}\right) \quad (30)$$

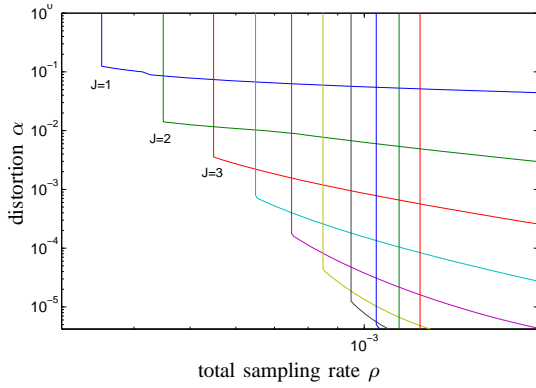


Fig. 5. The upper bound (Theorem 1) on the total sampling rate $\rho = Jr$ of the nearest subspace estimator as a function of the distortion α for various J when $\text{SNR} = 40$ dB and $\kappa = 10^{-4}$.

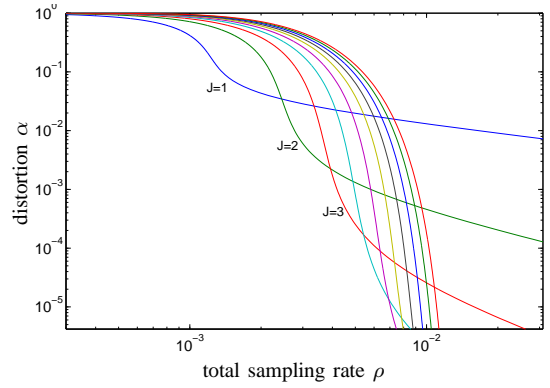


Fig. 6. The upper bound (Theorem 3) on the total sampling rate $\rho = Jr$ of LASSO + Joint Thresholding as a function of the distortion α for various J when $\text{SNR} = 30$ dB and $\kappa = 10^{-4}$.

for all α small enough.

Theorems 4 and 5 illustrate a tradeoff. At high SNR, the difficulty of estimation is dominated by the uncertainty about the nonzero values. Accordingly, the number of measurements is minimized by letting $J = 1$. As the desired distortion becomes small however, the opposite behavior occurs. Since estimation is limited by the size of the smallest nonzero values, it is optimal to choose J large to increase the diversity power. This behavior can be seen, for example, in Figures 3-6.

A natural question then, is how does one best choose the diversity J ? The following result shows that the right amount of diversity can significantly improve performance.

Theorem 6. *Let (κ, SNR) be fixed and let $\rho(\alpha, J)$ denote the infimum over sampling rates ρ such that α is achievable with diversity J . Then,*

$$\rho(\alpha, J) \leq \kappa J + O\left(\frac{\alpha}{P(\alpha, J)}\right). \quad (31)$$

Moreover, if $J = J^*(\alpha) = \Theta(\log(1/\alpha))$ then

$$\rho(\alpha, J^*(\alpha)) = \Theta(\log(1/\alpha)). \quad (32)$$

An important implication of Theorem 6 is that the optimal choice of J allows the distortion to decay *exponentially* rapidly with the sampling rate ρ . Note that the rate of decay is only polynomial if J is fixed. Interestingly, it can also be shown that the same exponential boost can be obtained using non-optimal estimators, albeit with smaller constants in the exponent.

The effect of the diversity J is illustrated in Fig. 5 for the nearest subspace estimator and in Fig. 6 for Lasso + thresholding. In both cases, the bounds show the same qualitative behavior—each value of the diversity J traces out a different curve in the sampling rate distortion region. It is important to note however, that due to the sub-optimality of the two stage architecture and the LASSO estimator, these similar behaviors occur only at different SNRs and with an order of magnitude difference in the sampling rate.

ACKNOWLEDGMENT

This work was supported in part by ARO MURI No. W911NF-06-1-0076, and in part by 3TU.CeDICT: Centre for Dependable ICT Systems, The Netherlands.

REFERENCES

- [1] M. J. Wainwright, "Information-theoretic limitations on sparsity recovery in the high-dimensional and noisy setting," *IEEE Trans. Inform. Theory*, vol. 55, pp. 5728–5741, Dec. 2009.
- [2] G. Reeves and M. Gastpar, "Approximate sparsity pattern recovery: Information-theoretic lower bounds," Feb. 2010, arXiv:1002.4458v1 [cs.IT].
- [3] —, "Fundamental tradeoffs for sparsity pattern recovery," Jun. 2010, arXiv:1006.3128v1.
- [4] D. Baron, M. F. Duarte, M. B. Wakin, S. Sarvotham, and R. G. Baraniuk, "Distributed compressed sensing," Rice University, Department of Electrical and Computer Engineering, Tech. Rep. TREE-0612, Nov. 2006.
- [5] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde, "Model-based Compressive Sensing," *IEEE Trans. Inform. Theory*, vol. 56, pp. 1982–2001, Apr. 2010.
- [6] N. Vaswani and W. Lu, "Modified-cs: Modifying compressive sensing for problems with partially known support," *IEEE Trans. Signal Process.*, vol. 58, no. 9, pp. 4595–4607, Sep. 2010.
- [7] S. F. Cotter, B. D. Rao, K. Engan, and K. Kreutz-delgado, "Sparse solutions to linear inverse problems with multiple measurement vectors," *IEEE Trans. Signal Process.*, vol. 53, no. 7, pp. 2477–2488, Jul. 2005.
- [8] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, "Algorithms for simultaneous sparse approximation part I: Greedy pursuit," *Signal Processing*, vol. 86, pp. 572–588, Apr. 2006.
- [9] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *J. of the Royal Stat. Soc. Series B*, vol. 68, no. 1, pp. 49–67, 2006.
- [10] G. Obozinski, M. J. Wainwright, and M. I. Jordan, "Union support recovery in high-dimensional multivariate regression," Department of Statistics, UC Berkeley, Tech. Rep., Aug. 2008.
- [11] K. Lounici, M. Pontil, A. B. Tsybakov, and S. van de Geer, "Taking advantage of sparsity in multi-task learning," arXiv:0903.1468v1 [stat.ML], Mar. 2009.
- [12] Y. C. Eldar, P. Kuppinger, and H. Bolcskei, "Block-sparse signals: Uncertainty relations and efficient recovery," *IEEE Trans. Signal Process.*, vol. 58, no. 6, pp. 3042–3054, Jun. 2010.
- [13] M. Stojnic, F. Parvaresh, and B. Hassibi, "On the reconstruction of block-sparse signals with an optimal number of measurements," *IEEE Trans. Signal Process.*, vol. 57, no. 8, pp. 3075–3085, Aug. 2009.
- [14] G. Reeves and M. Gastpar, "Efficient sparsity pattern recovery," in *Proc. 30th Symposium on Information Theory*, Eindhoven, May. 2009.
- [15] M. Bayati and A. Montanari, "The dynamics of message passing on dense graphs, with applications to compressed sensing," Mar. 2010, arXiv:1001.3448v2 [cs.IT].
- [16] D. L. Donoho, A. Maleki, and A. Montanari, "Message passing algorithms for compressed sensing," in *Proc. National Academy of Sciences*, vol. 106, 2009, pp. 18 914–18 919.
- [17] D. Guo and S. Verdú, "Randomly spread cdma: Asymptotics via statistical physics," *IEEE Trans. Inform. Theory*, vol. 51, no. 6, pp. 1983–2010, Jun. 2005.